

Article

Genome-Wide Identification and Characterization of Trihelix Gene Family in Asian and African *Vigna* Species

Shweta Kumari ¹, Dhammaprakash Pandhari Wankhede ^{2,*} , Sneha Murmu ¹ , Ranjeet Maurya ^{2,3} , Sarika Jaiswal ¹ , Anil Rai ¹  and Sunil Archak ^{2,*} 

¹ Division of Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute (IASRI), Pusa Campus, New Delhi 110012, India

² ICAR-National Bureau Plant Genetic Resources, Pusa Campus, New Delhi 110012, India

³ CSIR-Institute of Genomics and Integrative Biology, New Delhi 110007, India

* Correspondence: d.wankhede@icar.gov.in (D.P.W.); sunil.archak@icar.gov.in (S.A.); Tel.: +91-11-25802875 (D.P.W.); +91-11-25802883 (S.A.)

Abstract: Trihelix transcription factors play a crucial role in varied stress responses as well as in the growth and development of plants. The role of trihelix transcription factors in the non-shattering phenotype in domesticated rice is known. The *Vigna* group of crops has different degrees of shattering phenotypes in different species. To understand the evolutionary conservation or divergence of the trihelix gene family in important *Vigna* species here, the genome-wide identification and characterization of the trihelix gene family in four *Vigna* species including the cowpea (*Vigna unguiculata*), mung bean (*V. radiata*), adzuki bean (*V. angularis*) and rice bean (*V. umbellata*) was performed. A total of 39, 35, 41 and 50 trihelix genes were identified in the cowpea, mung bean, adzuki bean and rice bean, respectively. The trihelix genes in each of the four *Vigna* species were classified into five subgroups: GT, GT γ , SH4, S1P1 and GT δ . The members of each subgroup shared similar patterns of gene structure and motif across the four species. The cross-species positional relationships of the cowpea, adzuki bean and mung bean vis-a-vis rice trihelix genes were studied. Further, the *Ka/Ks* ratio for the trihelix genes in the four *Vigna* species indicated the purifying or stabilizing selection of the family. The gene expression analysis of the trihelix gene family in the cowpea showed that most of the genes express in at least some of the seed and/or pod developmental stages, although at varying degrees. Based on detailed bioinformatic analysis, a potential target for gene editing towards a possible non-shattering phenotype in the four important *Vigna* crops was discussed.

Keywords: trihelix transcription factor; gene family; in silico characterization; domestication; pod dehiscence



Citation: Kumari, S.; Wankhede, D.P.; Murmu, S.; Maurya, R.; Jaiswal, S.; Rai, A.; Archak, S. Genome-Wide Identification and Characterization of Trihelix Gene Family in Asian and African *Vigna* Species. *Agriculture* **2022**, *12*, 2172. <https://doi.org/10.3390/agriculture12122172>

Academic Editor: Gaoneng Shao

Received: 16 November 2022

Accepted: 7 December 2022

Published: 18 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The domestication of plants is one of the most crucial developments in the history of agriculture. Domesticated plants differ from their wild predecessors in several characteristics and the non-shattering of seeds or pod indehiscence is one of the most crucial traits [1,2]. Seed shattering or pod dehiscence is the detachment of the seed naturally at maturity, which leads to inefficient harvesting and substantial yield loss [3]. In several species, the indehiscent phenotype is acquired independently through convergent evolution leading to the same functional changes [4,5]. In rice (*Oryza sativa*), *sh4* has been identified as a major QTL on chromosome 4 responsible for the reduction of shattering in cultivated rice [6]. Further, fine mapping and validation have revealed that a single mutation causing the change in nucleotide guanine (G) to thymine (T) leading to the substitution of the amino acid lysine (K) to asparagine (N) in the transcription factor at *Sh4* loci which has been shown to be responsible for the non-shattering phenotype in cultivated rice [7]. In addition to this, a

single dominant gene, 'Shattering 1' (*SHA1*), was identified from perennial wild rice (*O. rufipogon*) that controls seed shattering. *SHA1* also harbored the same amino acid 'K79' as that of the wild rice *O. nivara* at *sh4* locus. *SHA1* was found allelic to *sh4* and encodes a trihelix family transcription factor (TF) [8].

The trihelix family is one amongst the earliest discovered transcription factors in plants. Initially, the trihelix transcription factor was classified as a GT factor because of its binding specificity for the light-responsive GT element. The DNA-binding domain of the GT factor constitutes a typical helix (helix-loop-helix-loop-helix) structure with a degenerate core sequence of 5'-G-Pu-(T/A)-A-A-(T/A)-3' [9,10]. It was observed that the structures of the GT factor trihelix domains and Myb DNA-binding domains were moderately similar [11]; however, they have different recognition sequences. Accordingly, a few databases have even classified the trihelix proteins as Myb/SANT-LIKE domains (PF13837) containing transcription factors [10,11]. The first ever subfamily of trihelix TF discovered was GT-1, isolated from the pea (*Pisum sativum*), where it was found regulating the light-dependent expression of the *rbcS-3A* gene by binding on the promoter region. The expression of GT-1 is light-inducible with core sequence 5'-GGTTAA-3, after which the transcription factor was named [12,13]. The gene sequences that encode a DNA-binding protein with specificity for the GT-1 factor were isolated from tobacco (*Nicotiana tabacum*), and its predicted structure indicated the presence of three α -helices separated by a turn (HHTH) [14]. Further, a gene with another DNA-binding domain showed a similar conserved core sequence in the upstream region and binding preferences as that of the GT-1 factor, so it was named GT-2 [15,16]. The predicted structural characteristics of the GT-2 polypeptide segment were found as a helix-loop-helix-loop-helix motif, so it was designated as trihelix [17].

Recently, several trihelix genes were studied from various plants such as tomato, Arabidopsis, chrysanthemum and rice and were shown to have a role in different biological processes [18–21]. These functions include the regulation of light-dependent gene expression, responses to biotic and abiotic stress and a variety of developmental processes such as morphogenesis, seed shattering/non-shattering, stomatal development, late embryogenesis, seed development and responses to abiotic stress including salt and drought [8,10,22–25].

The genus 'Vigna' comprises important cultivated species such as the cowpea (*V. unguiculata* (L.) Walp), mung bean (*V. radiata* (L.) Wilczek), urd bean (*V. mungo* (L.) Hepper), azuki bean (*V. angularis* (Willd.) Ohwi & Ohashi) and rice bean (*V. umbellata* (Thunb.) Ohwi & Ohashi) with substantial economic and environmental importance. These species play a crucial role in nutritional security in tropical Asian and African countries [26]. Although domesticated, these cultivated species possess variable degrees of wild traits such as seed shattering, pre-harvest sprouting, etc.

Since the role of the trihelix transcription factor is established in conferring non-shattering traits in cultivated rice (8), it will be intriguing to know how far the trihelix gene family is conserved or diverged in *Vigna* species compared to rice and to what extent the parallel can be drawn from rice with respect to the shattering phenomenon. So far, there has been no detailed study of the trihelix gene family in *Vigna* species. The availability of the genomic resources in the form of the genome and transcriptome sequences in the *Vigna* group of crops [27–30] has opened opportunities to undertake genome-wide studies in these crops and to draw benefit from the comparative genome analysis. In the present work, the trihelix transcription factor gene family was identified and characterized in four *Vigna* species, cowpea, mung bean, adzuki bean and rice bean using bioinformatics and comparative genomic tools. The study of the gene families in these four crops simultaneously could help better understand cross-species genomic co-localization, conserved motif distribution, gene structure, sequence conservation and variation, functional attributes and evolutionary perspectives.

2. Materials and Methods

2.1. Genome-Wide Identification of Trihelix Genes in *Vigna* Species

The whole genome sequence, genes and protein sequences of cowpea, mung bean and adzuki bean were downloaded either from Ensembl Plants or Phytozome [31]. Rice bean transcriptome SRA data were downloaded from NCBI. The transcriptomes of *Vigna umbellata* were assembled by Trinity (v2.13.1; GitHub, San Francisco, CA, USA) [32]. The de novo transcriptome assembly was then used to remove redundant sequences by CD-HIT [33]. The *Vigna umbellata* gene prediction was implemented using the MAKER pipeline [34] and the output results were used for training the AUGUSTUS [35] model parameter for accuracy of gene prediction. Using the trained model parameter of *Vigna umbellata*, the prediction pipeline was again re-run against the repeat masked *Vigna umbellata* scaffolds. To identify the candidate trihelix family protein sequences in four *Vigna* crops, multiple sequence alignment was performed using T-coffee [36]. Signature domain of trihelix of different crops was downloaded from plant transcription factor databases. The HMM search (<https://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch>, accessed on 20 December 2019) was then performed with trihelix domain on whole genome protein sequence of each of the four crops. The significant hits with e-value <0.01 were then verified for the presence of trihelix domain using conserved domain search databases (CDD) [37] and Pfam [38]. The identified rice bean trihelix genes and protein sequences are given in Appendix A.

2.2. Phylogenetic Analysis

The multiple sequence alignment was carried out using ClustalW with amino acid sequences of trihelix family proteins in four *Vigna* crops and rice [21]. A phylogenetic tree was constructed based on above alignment by maximum likelihood (ML) method with the following parameters using MEGAX software: Poisson correction; pairwise deletion; 1000 bootstrap replicates [39]. Orthologous nomenclature was used for naming trihelix genes in four *Vigna* crops keeping rice trihelix genes as the basis.

2.3. Identification of Basic Features, Conserved Motifs and Gene Structure

The basic molecular features such as molecular weight, isoelectric point and coding sequence length of the trihelix proteins were obtained by using ProtParam tool [40]. The MEME-Suite web server [41] was used to identify the conserved motifs following default parameters. The trihelix gene sequences were extracted from genome file based on annotation attributes using custom-built Perl program. Exon/intron structure of trihelix genes was generated using Gene Structure Display Server [42]. Prediction of subcellular localization was obtained by using CELLO2GO [43].

2.4. Chromosomal Location and Evolutionary Selection Pressure

Genomic attributes of trihelix genes of cowpea, mung bean and adzuki bean were obtained from respective gene annotation files. Gene allocation on chromosomes was performed using MapChart (v.2.32; Wageningen University & Research, Wageningen, The Netherlands) software [44]. Circos [45] software was used to visualize the collinearity of orthologous trihelix genes of *Vigna* crops and rice. Evolutionary selection pressure was assessed by calculating K_a and K_s values for trihelix genes of each *Vigna* crop with its rice orthologs using PAL2NAL [46].

2.5. Expression Analysis of Trihelix Genes in Cowpea

Cowpea transcriptome SRA data (Table S1) [47] were used to study the expression pattern of trihelix genes in different tissue types (leaf, root, stem, pod, seed and mix-tissue) and time kinetics using RNA-Seq data analysis as per Trapnell et al. [48]. The 3-D structure of trihelix proteins with and without single amino acid change was constructed using I-TASSER server [49]. Minimum potential energy calculation for both protein versions (orig-

inal and mutated) was carried out on Gromacs (version 2020.1-Ubuntu-2020.1-1; University of Groningen, Groningen, The Netherlands) [50].

3. Results

3.1. Genome-Wide Identification, Phylogenetic Analysis and Nomenclature

The trihelix gene family in the cowpea, mung bean, rice bean and adzuki bean were identified using HMM search and BlastP followed by validation for the presence of the complete trihelix domain using the Pfam and InterPro databases. In the cowpea, mung bean, adzuki bean and rice bean, a total of 39, 35, 41 and 50 trihelix genes were identified, respectively. The structural and functional attributes of the trihelix family members for the cowpea, mung bean, adzuki bean and rice bean are given in Tables 1–4, respectively. To understand the relatedness of trihelix genes among the four *Vigna* groups of crops and with rice, phylogenetic analysis was performed (Figure 1). All the trihelix genes of the four *Vigna* crops were clustered in three major groups. As per the classification of trihelix genes in rice and the orthologous relationship, five subgroups, were assigned as GT, GT δ , GT γ , S1P1 and SH4 to the identified trihelix genes in the four *Vigna* crops. The GT γ was found to be the largest subgroup in all four *Vigna* species, comprising 18, 16, 21 and 25 genes in the cowpea, mung bean, adzuki bean and rice bean, respectively. The second largest subgroup was S1P1, which consisted of eight members each in the cowpea and adzuki bean, while there were four and ten members in the mung bean and rice bean, respectively. GT δ was the smallest subgroup, comprising one member in the cowpea and adzuki bean and three members each in the mung bean and rice bean. The SH4 and GT subgroups showed nine and three members in the cowpea, six members each in the mung bean, seven and four members in the adzuki bean and seven and five members in the rice bean, respectively. Most of the members of the subfamily in the trihelix in all four *Vigna* and rice were clustered together, albeit with a few exceptions.

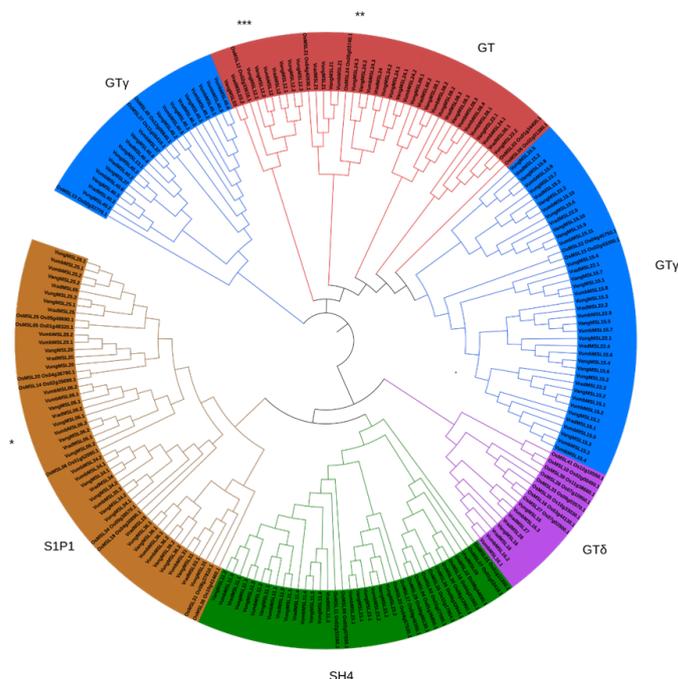


Figure 1. Phylogenetic tree based on protein sequence of trihelix genes in cowpea, mung bean, adzuki bean, rice bean and rice. The maximum likelihood tree was created using MEGAX (bootstrap value = 1000). Vung, *V. unguiculata*; Vrad, *V. radiata*; Vang, *V. angularis*; Vumb, *V. umbellata*; Os, *Oryza sativa*. MSL, Myb/SANT-LIKE. The clade in red represents GT family, blue represents GT γ family, brown represents S1P1 family, green represents SH4 family and magenta represents GT δ family. ‘*’, ‘**’ and ‘***’ represent member of GT, GT γ and SH4 family clustered elsewhere.

Table 1. Basic characteristics of Cowpea (*Vigna unguiculata*) Trihelix genes.

Subgroup	Assigned Name	Locus ID	Exon (No.)	Amino Acid (No.)	^a MW (kDa)	^b PI	GC%	^c Loc.	Rice Orthologs
GT	VungMSL08.1	Vigun06g111200	2	569	64.21	6.14	42.6	N	OsMSL08 (Os02g01380.1)
	VungMSL08.2	Vigun07g154500	2	617	70.95	6.78	40.9	N	
	VungMSL21	Vigun01g244400	4	377	43.14	5.83	37.6	N, C	OsMSL21 (Os04g40930.1)
SH4	VungMSL11.1	Vigun05g020800	3	326	36.07	5.23	38	N	OsMSL11 (Os02g31160.1)
	VungMSL11.2	Vigun07g298000	2	349	38.86	6.5	37.4	N	
	VungMSL11.3	Vigun07g298000	2	260	29.00	6.03	37.4	N	
	VungMSL11.4	Vigun03g052500	2	339	38.02	6.87	41.9	N	OsMSL12 (Os02g33610.1)
	VungMSL12.1	Vigun03g314600	17	883	98.29	8.76	33.7	Ch	
	VungMSL12.2	Vigun09g101800	17	723	79.93	8.64	35.5	M	
	VungMSL12.3	Vigun09g101800	15	723	79.93	8.64	35.5	M	OsMSL23 (Os04g57530.1)
	VungMSL23.1	Vigun07g196800	2	346	39.75	9.03	43.6	N	
	VungMSL23.2	Vigun02g002400	2	344	39.14	7.73	32.8	N	
	S1P1	VungMSL20	Vigun10g160800	1	376	40.79	9.73	47.4	N
VungMSL25.1		Vigun06g191000	2	301	33.97	5.15	39	N	OsMSL25 (Os05g48690.1)
VungMSL25.2		Vigun05g208200	2	300	33.84	9.76	35.7	N	
VungMSL31		Vigun10g048800	1	262	30.16	6.68	40.7	N	OsMSL31 (Os08g37810.1)
VungMSL34.1		Vigun11g011300	2	339	38.16	9.25	30.5	N	OsMSL34 (Os09g38570.1)
VungMSL34.2		Vigun01g027600	2	349	38.86	8.9	38.2	N	
VungMSL36.1		Vigun05g050600	3	245	27.32	10.47	61.3	N, M	OsMSL36 (Os10g41460.1)
VungMSL36.2		Vigun03g429900	1	310	34.73	9.34	49.3	N	
GTδ	VungMSL16	Vigun06g021100	5	496	57.10	7.99	35.6	N, C	OsMSL16 (Os03g44130.1)
GTγ	VungMSL06.1	Vigun04g017300	1	352	40.73	4.62	39.5	N	OsMSL06 (Os01g52090.1)
	VungMSL06.2	Vigun05g286700	1	390	44.55	4.7	39.3	N	
	VungMSL13	Vigun07g065400	1	474	54.34	6.12	44	N	OsMSL13 (Os02g33770.1)
	VungMSL15.1	Vigun07g217300	2	655	72.32	5.94	44.7	N	OsMSL15 (Os02g43300.1)
	VungMSL15.2	Vigun04g131900	2	661	73.01	5.3	41.5	N	
	VungMSL15.3	Vigun04g131800	2	584	66.44	5.61	38.4	N	
	VungMSL15.4	Vigun09g125900	2	503	56.99	6.44	43.6	N	
	VungMSL15.5	Vigun07g193200	2	519	59.5	5.76	41.3	N	
	VungMSL15.6	Vigun01g170400	2	586	67.6	6.38	41.4	N	OsMSL22 (Os04g45750.1)
	VungMSL15.7	Vigun07g067700	2	602	69.77	6.32	41.7	N	
	VungMSL22.1	Vigun07g217100	2	815	92.69	5.24	41.3	N	
	VungMSL22.2	Vigun06g151000	2	268	31.34	7.63	35.5	N	
	VungMSL24.1	Vigun10g080500	2	274	32.528	6.65	38.5	N	OsMSL24 (Os05g03740.1)
	VungMSL24.2	Vigun05g100400	2	289	34.71	6.99	32.9	N	
	VungMSL24.3	Vigun02g104900	2	309	36.26	7.74	39	N	
VungMSL40.1	Vigun07g217400	1	450	51.34	6.2	38.8	N	OsMSL40 (Os12g06640.1)	
VungMSL40.2	Vigun05g263000	1	447	50.73	5.78	39.2	N		
VungMSL40.3	Vigun01g168100	1	429	49.48	6.22	44.7	N		

^a molecular weight ^b isoelectric point ^c subcellular localization.

Table 2. Basic characteristics of *Vigna radiata* Trihelix genes.

Sub-group	Assigned Name	Locus ID	Exon (No.)	Amino Acid (No.)	^a MW (kDa)	^b PI	GC%	^c Loc.	Rice Orthologs	
GT	<i>VradMSL03.1</i>	<i>Vradi09g07980.1</i>	2	203	23.45	7.69	42.5	N	<i>OsMSL03</i> (<i>Os01g34400.1</i>)	
	<i>VradMSL03.2</i>	<i>Vradi01g07000.1</i>	5	987	110.71	7.1	45	N		
	<i>VradMSL08.1</i>	<i>Vradi10g07270.1</i>	3	282	32.00	7.77	41.8	N	<i>OsMSL08</i> (<i>Os02g01380.1</i>)	
	<i>VradMSL08.2</i>	<i>Vradi10g05520.1</i>	3	556	62.56	8.18	43.9	N		
	<i>VradMSL08.3</i>	<i>Vradi08g13180.1</i>	2	590	67.75	7	42.9	N		
	<i>VradMSL21</i>	<i>Vradi02g12120.1</i>	7	396	45.15	5.9	49.37	N, C	<i>OsMSL21</i> (<i>Os04g40930.1</i>)	
SH4	<i>VradMSL09</i>	<i>Vradi06g09390.1</i>	5	451	51.29	5.2	37.8	N	<i>OsMSL09</i> (<i>Os02g07800.1</i>)	
	<i>VradMSL11.1</i>	<i>Vradi08g23420.1</i>	3	231	26.23	9	35.4	N	<i>OsMSL11</i> (<i>Os02g31160.1</i>)	
	<i>VradMSL11.2</i>	<i>Vradi04g10090.1</i>	3	326	35.99	5.24	38.4	N		
	<i>VradMSL11.3</i>	<i>Vradi07g29850.1</i>	3	300	34.51	9.3	41.4	N		
		<i>VradMSL11.4</i>	<i>Vradi07g01760.1</i>	4	227	26.20	9.2	30.9	N	
	<i>VradMSL12</i>	<i>Vradi0043s00060</i>	15	873	96.84	8.5	35.6	M	<i>OsMSL12</i> (<i>Os02g33610.1</i>)	
S1P1	<i>VradMSL05</i>	<i>Vradi10g10130.1</i>	4	258	29.10	10.06	33.1	N, M	<i>OsMSL05</i> (<i>Os01g48320.1</i>)	
	<i>VradMSL20</i>	<i>Vradi09g02750.1</i>	3	239	26.12	9.98	43.7	N	<i>OsMSL20</i> (<i>Os04g36790.1</i>)	
	<i>VradMSL25</i>	<i>Vradi0023s00720.1</i>	2	212	24.22	9.85	59.17	N	<i>OsMSL25</i> (<i>Os05g48690.1</i>)	
	<i>VradMSL34</i>	<i>Vradi05g17220.1</i>	3	149	16.78	9.86	33.2	N	<i>OsMSL34</i> (<i>Os09g38570.1</i>)	
GTδ	<i>VradMSL16</i>	<i>Vradi05g11080.1</i>	7	877	101.45	5.4	39.48	N	<i>OsMSL16</i> (<i>Os03g44130.1</i>)	
	<i>VradMSL27</i>	<i>Vradi07g00070.1</i>	6	762	87.92	6.25	40.37	N, C	<i>OsMSL27</i> (<i>Os07g02500.1</i>)	
	<i>VradMSL28</i>	<i>Vradi03g06280.1</i>	4	482	55.85	6.49	39.61	N	<i>OsMSL28</i> (<i>Os07g10950.1</i>)	
GTγ	<i>VradMSL06.1</i>	<i>Vradi05g20210.1</i>	2	278	31.72	5.36	41.34	N	<i>OsMSL06</i> (<i>Os01g52090.1</i>)	
	<i>VradMSL06.2</i>	<i>Vradi01g04230.1</i>	3	242	27.80	4.8	51.17	N, C		
	<i>VradMSL15.1</i>	<i>Vradi08g18150.1</i>	2	445	49.91	9.81	46.52	N	<i>OsMSL15</i> (<i>Os02g43300.1</i>)	
	<i>VradMSL15.2</i>	<i>Vradi08g16250.1</i>	2	518	59.58	5.78	45.54	N		
	<i>VradMSL15.3</i>	<i>Vradi08g06550.1</i>	3	497	58138	6.18	43.91	N		
		<i>VradMSL19</i>	<i>Vradi08g05950.1</i>	6	485	53.11	5.87	48.9	N, Ch	<i>OsMSL19</i> (<i>Os04g30890.1</i>)
		<i>VradMSL22.1</i>	<i>Vradi0160s00310.1</i>	5	307	35.26	9.17	49.78	N	<i>OsMSL22</i> (<i>Os04g45750.1</i>)
		<i>VradMSL22.2</i>	<i>Vradi01g12370.1</i>	4	446	51.30	9.3	44.82	N	
		<i>VradMSL22.3</i>	<i>Vradi01g12360.1</i>	4	277	32147	4.93	50.84	N, C	
		<i>VradMSL22.4</i>	<i>Vradi08g18140.1</i>	2	367	41.80	9.44	45.56	N	
	<i>VradMSL22.5</i>	<i>Vradi03g04320.1</i>	2	590	67.77	6.29	43.36	N		
	<i>VradMSL24</i>	<i>Vradi04g00170.1</i>	2	233	28.78	9.57	46.87	N, M	<i>OsMSL24</i> (<i>Os05g03740.1</i>)	
	<i>VradMSL40.1</i>	<i>Vradi08g18160.1</i>	3	351	40.07	5.49	42.05	N	<i>OsMSL40</i> (<i>Os12g06640.1</i>)	
	<i>VradMSL40.2</i>	<i>Vradi03g04100.1</i>	1	225	25.54	8.89	46.9	N		
	<i>VradMSL40.3</i>	<i>Vradi04g01660.1</i>	2	341	38.41	8.93	43.47	N		
	<i>VradMSL40.4</i>	<i>Vradi08g05970.1</i>	2	341	39.88	9.59	46.2	N		

^a molecular weight ^b isoelectric point ^c subcellular localization. N- Nucleus, C-Cytoplasm, Ch- Chloroplast.

Table 3. Basic Characteristics of *Vigna angularis* Trihelix genes.

Subgroup	Assigned Name	Locus ID	Exon (No.)	Amino Acid (No.)	^a MW (kDa)	^b PI	GC (%)	^c Loc.	Rice Orthologs
GT	VangMSL03	Vigan10g080900	4	1005	112.7	8	51.0	N	OsMSL03 (Os01g34400.1)
	VangMSL08.1	Vigan09g161100	2	582	65.57	6.1	45.7	N	OsMSL08 (Os02g01380.1)
	VangMSL08.2	Vigan10g097100	2	588	67.28	6.5	46.5	N	
	VangMSL21	Vigan03g313100	5	392	44.86	5.9	49.1	N, C	OsMSL21 (Os04g40930.1)
SH4	VangMSL11.1	Vigan03g008100	3	322	35.66	5.4	54.4	N	OsMSL11 (Os02g31160.1)
	VangMSL11.2	Vigan02g258700	2	357	39.58	6.1	39.3	N	
	VangMSL11.3	Vigan01g003800	3	359	40.52	5.6	48.3	N	
	VangMSL12.1	Vigan1112s00030016		849	94.21	8.9	42.2	M	OsMSL12 (Os02g33610.1)
	VangMSL12.2	Vigan01g226300	2	167	18.56	5.3	38.8	N	
	VangMSL23.1	Vigan02g183900	2	346	39.30	9.2	52.8	N	OsMSL23 (Os04g57530.1)
	VangMSL23.2	Vigan06g146100	2	344	39.14	8.6	49.2	N	
S1P1	VangMSL20	Vigan11g006000	1	374	40.64	9.7	56.8	N	OsMSL20 (Os04g36790.1)
	VangMSL25.1	Vigan05g139300	1	582	65.57	6.1	55.7	N	OsMSL25 (Os05g48690.1)
	VangMSL25.2	Vigan09g234100	1	330	37.53	5.1	55.7	N	
	VangMSL31	Vigan11g133500	1	261	29.99	6.3	57.6	N	OsMSL31 (Os08g37810.1)
	VangMSL34.1	Vigan05g004000	2	349	38.83	9.2	45.9	N	OsMSL34 (Os09g38570.1)
	VangMSL34.2	Vigan08g003100	2	325	36.90	8.9	47.6	N	
	VangMSL36.1	Vigan03g034300	1	317	35.20	9.5	61.6	N	OsMSL36 (Os10g41460.1)
	VangMSL36.2	Vigan01g314100	1	309	34.65	9.3	56.2	N	
GTδ	VangMSL16	Vigan04g227300	7	829	96.05	5.3	39.7	N	OsMSL16 (Os03g44130.1)
GTγ	VangMSL06.1	Vigan10g017200	1	349	40.51	4.6	52.5	N	OsMSL06 (Os01g52090.1)
	VangMSL06.2	Vigan05g214400	1	387	44.21	4.7	41.8	N	
	VangMSL15.1	Vigan10g174500	2	120	13.59	8.7	44.9	N	
	VangMSL15.2	Vigan10g132200	2	657	72.52	5.3	50.7	N	
	VangMSL15.3	Vigan02g203500	3	682	75.06	6.0	50.4	N	
	VangMSL15.4	Vigan02g203300	2	489	54.30	9	45.9	N	OsMSL15 (Os02g43300.1)
	VangMSL15.5	Vigan10g132400	3	625	71.03	5.5	45.2	N	
	VangMSL15.6	Vigan02g203400	3	605	68.21	5.6	46	N	
	VangMSL15.7	Vigan04g134500	2	514	57.98	6.6	49.6	N	
	VangMSL15.8	Vigan02g180100	2	516	59.46	5.7	45.3	N	
	VangMSL15.9	Vigan03g238700	2	325	36.68	8.4	48.5	N	
	VangMSL15.10	Vigan03g238600	2	557	63.51	6.3	46.8	N	
	VangMSL22.1	Vigan09g201300	2	329	37.71	6.3	47.2	N	OsMSL22 (Os04g45750.1)
	VangMSL22.2	Vigan02g080400	2	553	63.51	5.8	45.8	N	
VangMSL24.1	Vigan11g113800	2	270	32.15	6.6	47.3	N	OsMSL24 (Os05g03740.1)	
VangMSL24.2	Vigan06g074900	2	309	36.33	8.3	49.3	N		

Table 3. Cont.

Subgroup	Assigned Name	Locus ID	Exon (No.)	Amino Acid (No.)	^a MW (kDa)	^b PI	GC (%)	^c Loc.	Rice Orthologs
	VangMSL37	Vigan641s002000	1	410	47.62	6.4	46.5	N	OsMSL37 (Os11g06410.1)
	VangMSL40.1	Vigan02g203600	1	448	51.27	6.2	42.2	N	
	VangMSL40.2	Vigan10g131700	1	442	50.57	6.2	42.5	N	OsMSL40
	VangMSL40.3	Vigan05g194200	1	448	51.02	5.9	42.6	N	(Os12g06640.1)
	VangMSL40.4	Vigan02g076800	1	430	49.57	6.1	46.9	N	

^a molecular weight ^b isoelectric point ^c subcellular localization. N- Nucleus, C-Cytoplasm.

Table 4. Basic Characteristics of *Vigna umbellata* Trihelix genes.

Subgroup	Assigned Name	Locus ID	Length	^a MW	^b PI	GC%	^c Loc.	Rice Orthologs
GT	VumbMSL08.1	Gene_28670	424	49.13	7.24	46.4	N	
	VumbMSL08.2	Gene_13448	582	65.56	6.18	45.7	N	OsMSL08
	VumbMSL08.3	Gene_13439	588	67.37	6.58	46.3	N	(Os02g01380.1)
	VumbMSL08.4	Gene_13440	459	52.32	7.77	45.6	N	
	VumbMSL21	Gene_22669	377	43.15	5.83	49.4	N, C	OsMSL21 (Os04g40930.1)
SH4	VumbMSL11.1	Gene_1713	210	23.08	5.40	60.5	N	
	VumbMSL11.2	Gene_1714	322	35.66	5.42	54.6	N	OsMSL11
	VumbMSL11.3	Gene_5885	363	40.07	6.10	50.4	N	(Os02g31160.1)
	VumbMSL11.4	Gene_30530	285	32.25	5.45	48	N	
	VumbMSL12	Gene_21287	867	96.37	8.76	41.8	Ch	OsMSL12 (Os02g33610.1)
	VumbMSL23.1	Gene_32537	340	38.61	9.03	53.2	N	
	VumbMSL23.2	Gene_29588	340	38.77	8.85	49.5	N	OsMSL23 (Os04g57530.1)
	VumbMSL20.1	Gene_23931	374	40.64	9.73	56.8	N	OsMSL20
	VumbMSL20.2	Gene_23932	374	40.64	9.73	56.8	N	(Os04g36790.1)
	VumbMSL25.1	Gene_14426	301	33.97	5.15	57.6	N	OsMSL25
	VumbMSL25.2	Gene_14425	330	37.59	5.17	55.6	N	(Os05g48690.1)
S1P1	VumbMSL31	Gene_2041	261	30.14	6.62	57.5	N	OsMSL31 (Os08g37810.1)
	VumbMSL34.1	Gene_18487	206	22.51	9.16	49.5	N	
	VumbMSL34.2	Gene_18488	349	38.83	9.15	46.4	N	OsMSL34
	VumbMSL34.3	Gene_18489	333	37.76	9.02	47.5	N	(Os09g38570.1)
	VumbMSL36.1	Gene_3480	317	35.20	9.55	61.6	N	OsMSL36
	VumbMSL36.2	Gene_3479	309	34.68	9.34	56.6	N	(Os10g41460.1)
GTδ	VumbMSL16.1	Gene_26106	487	56.85	6.63	40.4	N	
	VumbMSL16.2	Gene_26107	469	548.75	6.97	40.3	N	OsMSL16
	VumbMSL16.3	Gene_28699	505	58.67	9.26	39.2	C	(Os03g44130.1)

^a molecular weight ^b isoelectric point ^c subcellular localization.

Table 4. Cont.

Subgroup	Assigned Name	Locus ID	Length	^a MW	^b PI	GC%	^c Loc.	Rice Orthologs
GT γ	VumbMSL06.1	Gene_25832	349	40.50	4.69	52.4	N	OsMSL06 (Os01g52090.1)
	VumbMSL06.2	Gene_25833	349	40.50	4.69	52.4	N	
	VumbMSL06.3	Gene_9047	387	44.21	4.68	41.7	N	
	VumbMSL15.1	Gene_16032	555	62.55	5.61	49.5	N	OsMSL15 (Os02g43300.1)
	VumbMSL15.2	Gene_16033	657	72.55	5.35	50.8	N	
	VumbMSL15.3	Gene_16028	549	61.88	6.25	48.6	N	
	VumbMSL15.4	Gene_16030	652	72.09	5.82	50.3	N	
	VumbMSL15.5	Gene_16029	130	15.13	4.43	52.7	N, C	
	VumbMSL15.6	Gene_16031	630	71.07	5.68	46.5	N	
	VumbMSL15.7	Gene_16026	601	68.34	5.69	45.5	N	
	VumbMSL15.8	Gene_24101	632	71.53	5.96	50	N	
	VumbMSL15.9	Gene_195	518	59.55	5.80	45.5	N	
	VumbMSL15.10	Gene_18868	607	70.29	6.33	46.5	N	
	VumbMSL15.11	Gene_18869	589	67.60	6.22	46.6	N	
	VumbMSL22	Gene_16025	123	14.53	4.26	42.9	N	OsMSL22 (Os04g45750.1)
	VumbMSL24.1	Gene_28671	256	29.88	7.66	42.4	N	OsMSL24 (Os05g03740.1)
	VumbMSL24.2	Gene_18886	269	31.96	6.39	47.4	N	
	VumbMSL24.3	Gene_10842	309	36.33	8.36	49.4	N	
	VumbMSL37	Gene_5137	149	17.21	9.55	45.9	N	OsMSL37 (Os11g06410.1)
	VumbMSL40.1	Gene_11435	448	51.27	6.20	42.3	N	OsMSL40 (Os12g06640.1)
	VumbMSL40.2	Gene_11436	448	51.27	6.20	42.3	N	
VumbMSL40.3	Gene_11437	448	51.27	6.20	42.3	N		
VumbMSL40.4	Gene_11438	448	51.27	6.20	42.3	N		
VumbMSL40.5	Gene_11439	442	50.58	6.17	42.7	N		
VumbMSL40.6	Gene_28728	448	50.99	5.94	42.5	N		

^a molecular weight ^b isoelectric point ^c subcellular localization.

In order to maintain uniformity of nomenclature in all four *Vigna* groups of crops, unified orthology-based nomenclature was followed keeping rice trihelix names as reference. In rice, the trihelix genes were named as OsMSL01 to OsMSL41 where MSL stands for 'Myb/SANT-LIKE'. This same naming pattern was extended to all four *Vigna* species. For the cowpea, the trihelix genes were named 'VungMSL' as a prefix for '*Vigna unguiculata* Myb/SANT-LIKE'; for the mung bean, 'VradMSL' for '*V. radiata* Myb/SANT-LIKE'; for the adzuki bean, VangMSL for '*V. angularis* Myb/SANT-LIKE'; and for the rice bean, 'VumbMSL' for '*V. umbellata* Myb/SANT-LIKE'. The crop-specific prefix was then followed by the gene number based on the orthologous gene number of the rice trihelix. For example, an orthologue of the rice trihelix gene OsMSL21 in the cowpea was named VungMSL21. Wherever multiple orthologous genes were present in *Vigna* species for one trihelix gene of rice, these genes were considered as paralogues in respective crops. The paralogues were numbered based on distance with respective rice orthologs and the least distance was named as '.1' and so on with increasing distance. For example, for OsMSL12 (Os02g33610.1), two genes were identified as orthologous in the cowpea and

were named based on their distance, with OsMSL12 as VungMSL12.1 (Vigun03g314600) and VungMSL12.2 (Vigun09g101800), respectively, (Figure 1, Tables 1–4).

3.2. Identification of Basic Features, Conserved Motifs, Gene Structure and Selection Pressure

The physical properties and other features, such as the length of CDS (coding sequence) and protein (amino acid), molecular weight, isoelectric point and subcellular localization, were predicted for the identified trihelix in all four *Vigna* crops (Tables 1–4). In the cowpea, the length of the coding sequence varied from 2655 bp (VungMSL12.1) to 738 bp (VungMSL36.1); in the adzuki bean this range was 3018 bp (VangMSL03) to 363 bp (VangMSL15.1). In the mung bean, the largest cds was 2964 bp (VradMSL03.2) and smallest one was 450 bp (VradMSL34), whereas in the rice bean this length differed from 2604 bp (VumbMSL12) to 373 bp (VumbMSL22). The predicted protein molecular weight of trihelix proteins ranged from 98.29 kDa (VungMSL12.1) to 27.32 kDa (VungMSL36.1) for the cowpea, 110.71 kDa (VradMSL03.2) to 16.78 kDa (VradMSL34) for the mung bean, 112.78 kDa (VangMSL03) to 13.92 kDa (VangMSL15.1) for the adzuki bean and 96.37 kDa (VumbMSL12) to 51.27 kDa (VumbMSL40.1) for the rice bean. Their predicted isoelectric points varied from 10.47 (VungMSL36.1) to 4.62 (VungMSL06.1) in the cowpea, 10.06 (VradMSL05) to 4.8 (VradMSL06.2) in the mung bean, 9.73 (VangMSL20) to 4.69 (VangMSL06.1) in the adzuki bean and 9.73 (VumbMSL20.1) to 4.26 (VumbMSL22) in the rice bean.

As per the prediction of sub-cellular localization, most of the trihelix proteins in the four *Vigna* species were found to have sub-cellular localization in the nucleus except a few. In the cowpea, the sub-cellular localization of VungMSL12.1 was found in the chloroplast, whereas in VungMSL12.2 and VungMSL12.3 in the mitochondria. In the mung bean, all the proteins were predicted to be localized in the nucleus except VradMSL12, which was found to be located in the mitochondria. In the adzuki bean, VangMSL12.1 was found to have localization in the mitochondria. In the rice bean, all the trihelix proteins except two showed nuclear localization. VumbMSL12 and VumbMSL16.3 were found to have chloroplastic and cytoplasmic localization, respectively.

To further analyze the evolutionary relationship, the exon–intron distribution of trihelix genes was studied in the cowpea, mung bean and adzuki bean along with rice (Figure 2), for which whole genome sequence with annotation is available in the public domain. The gene structure, in terms of exon numbers and exon–intron distribution, was mostly conserved for the orthologous genes of the rice and *Vigna* species, especially, among the *Vigna* species. However, there was observed a large variation in exon numbers and exon–intron distribution within the trihelix gene family for each of the four *Vigna* species. The number of exons in the trihelix family ranged from 1 to 17 in the cowpea, 2 to 15 in the mung bean and 1 to 16 in adzuki beans (Tables 1–4, Figure 2). Each subgroup of trihelix genes showed a specific pattern of exon number and exon–intron distribution. The average number of exons was lowest in the S1P1 family, which contained about one–three exons in the cowpea, two–four exons in the mung bean and one–two exons in the adzuki bean. The highest average number of exons was found in the SH4 group, which showed 2–17 exons in the cowpea, 2–16 exons in the adzuki bean and 3–15 exons in the mung bean. The second highest number of exons was in the GT δ group, and showed five exons in the cowpea, seven exons in the adzuki bean and four–seven exons in the mung bean. The trihelix subgroups, GT-1 and GT γ , showed two–four and one–two exons, respectively, in the cowpea; two–seven and one–six in the mung bean; two–five and one–three in the adzuki bean.

The non-synonymous to synonymous substitution rate (dN/dS) was calculated for every trihelix gene with respect to its rice ortholog. The calculated range of dN/dS rate varied from 0.0100 to 1.1412 in the GT subgroup, 0.0074 to 0.724 for GT γ , 0.0064 to 0.1020 for S1P1, 0.0039 to 0.086 for SH4 and from 0.0175 to 0.0782 in the GT δ subgroup (Tables S2–S5). The dN/dS value for all four *Vigna* trihelix genes with their rice ortholog pairs was found to be less than 1 except for VangMSL21, which indicated purifying or stabilizing

selection during the course of evolution. For VangMSL21, a member of the GT family, the dN/dS rate with respect to its rice ortholog OsMSL21 was 1.1412.

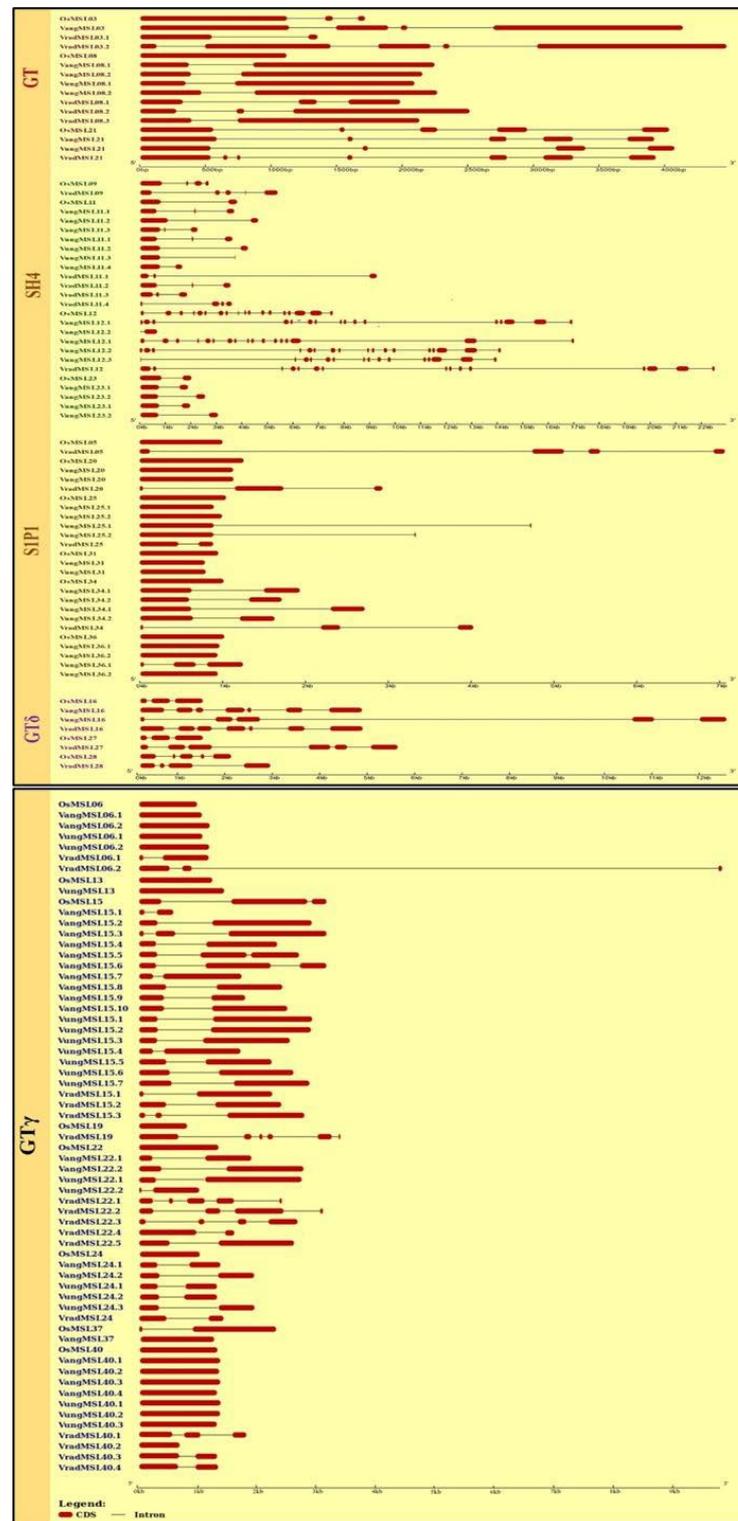


Figure 2. Gene structure of trihelix family genes in *Vigna* species along with its rice orthologs. The red bars represent exons and the black line between them represents introns. The scale at the bottom indicates the length of exons and introns. Vung, *V. unguiculata*; Vrad, *V. radiata*; Vang, *V. angularis*; Vumb, *V. umbellata*; Os, *Oryza sativa*. MSL, Myb/SANT-LIKE.

The conserved ‘helix-loop-helix’ domain of the trihelix protein was searched in all the trihelix proteins in *Vigna* by performing multiple sequence alignment of the Myb/SANT-LIKE domain. The well separated conserved motifs for three helices were observed in all the members of the five sub-families. Three individual α -helices with conserved amino acid tryptophan (W) were identified in GT, GT δ and SH4 (Figure S1, Figure S3 and Figure 4). In the S1P1 clade, other than the three individual α -helices with conserved tryptophan (W), there was an additional helix with conserved tryptophan followed by a continuous stretch of sequences (x-(F/Y)-(F/Y)-x-x-(L/M)-x-x-(L/M) (Figure S2). In the GT subfamily too, an additional helix with the conserved sequences ((Y/F)-(Y/Y)-x-x-(L/I/M)-x-x-(L/I/M)) was also found. The binding domain of the GT γ family comprised two individual α -helices with conserved tryptophan (W), whereas in third helix the conserved tryptophan (W) was found replaced by phenylalanine (F). The fourth helix with the sequence ((F/Y)-(F/Y)-x-x-(L/M)-x-x-(L/I)) was also found downstream of the third helix (Figure S4) in GT γ . This fourth helix was not observed in SH4; instead, they had the extended third helix with a conserved leucine (L). The same pattern was followed in all four *Vigna* species.

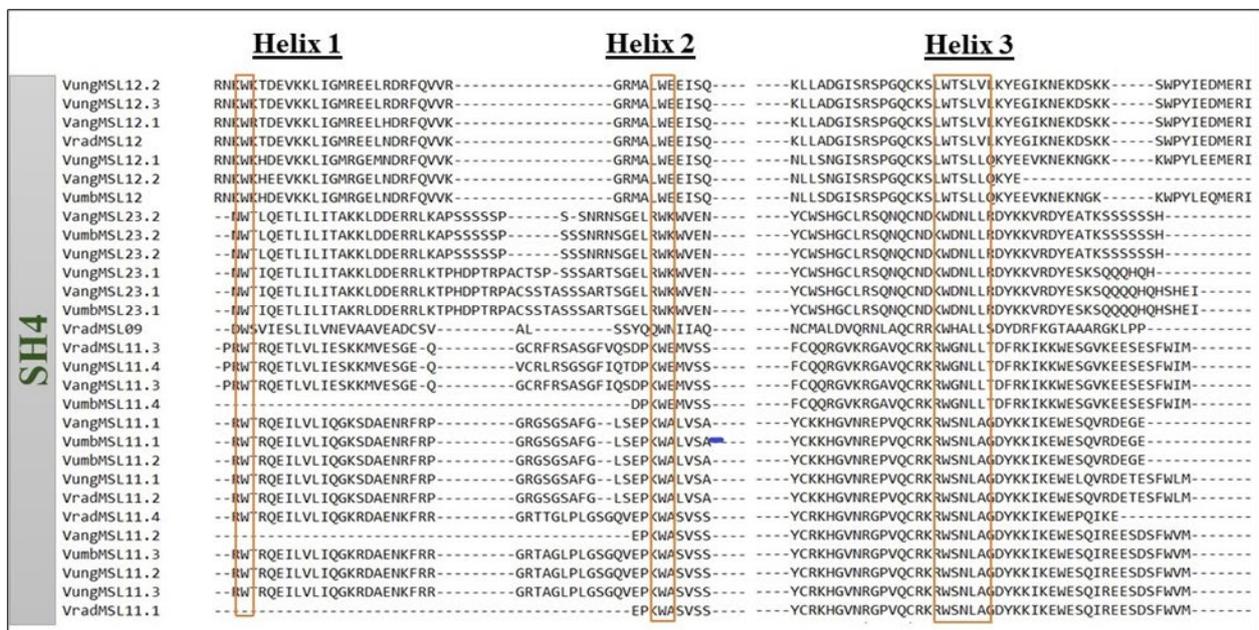


Figure 4. Sequence alignment of trihelix domain of SH4 group present in *Vigna* species. The three boxes represent the three helices of trihelix domain and sequence included in orange box represents conservation of tryptophan (W) present in the domain.

3.3. Chromosomal Distribution and Collinearity of Trihelix Genes among *Vigna* Species

The distribution of the trihelix family genes of the cowpea, mung bean and adzuki bean were mapped on chromosomes according to their loci position (Figure 5). For all three *Vigna* species, the genes were distributed on 10 of a total 11 chromosomes. In the cowpea, chromosome 7 harbored the highest number of trihelix genes, i.e., nine, whereas the least, two genes, were observed on both chromosome 2 and chromosome 9. For the adzuki bean, the highest (nine) and lowest (one) trihelix genes were observed on chromosomes 2 and 8, respectively. In the case of the mung bean, chromosome 8 comprised the largest (nine) number of genes, whereas chromosome 6 contained only one gene. No trihelix genes were located on chromosome 8 in the cowpea, chromosome 7 in the adzuki bean and chromosome 11 in the mung bean. The cross-species chromosomal positional relationship between trihelix genes in rice and their orthologs in three *Vigna* species are represented by a Circos plot (Figure 6). This also suggests a genomic rearrangement of the trihelix gene family in the studied *Vigna* species.

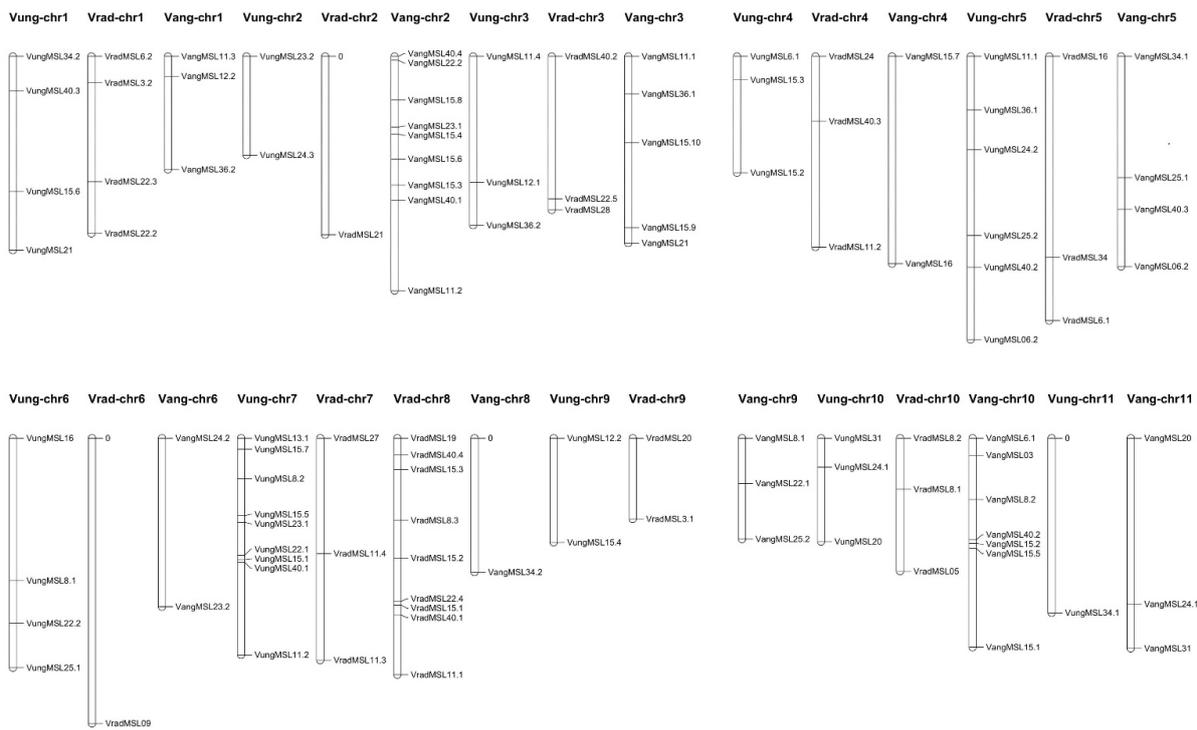


Figure 5. Chromosomal locations of trihelix genes of cowpea, mung bean and adzuki bean. Chromosome number is mentioned at the top. The chromosomes are presented in the order— cowpea (*Vung*, *V. unguiculata*), mung bean (*Vrad*, *V. radiata*) and then adzuki bean (*Vang*, *V. angularis*).

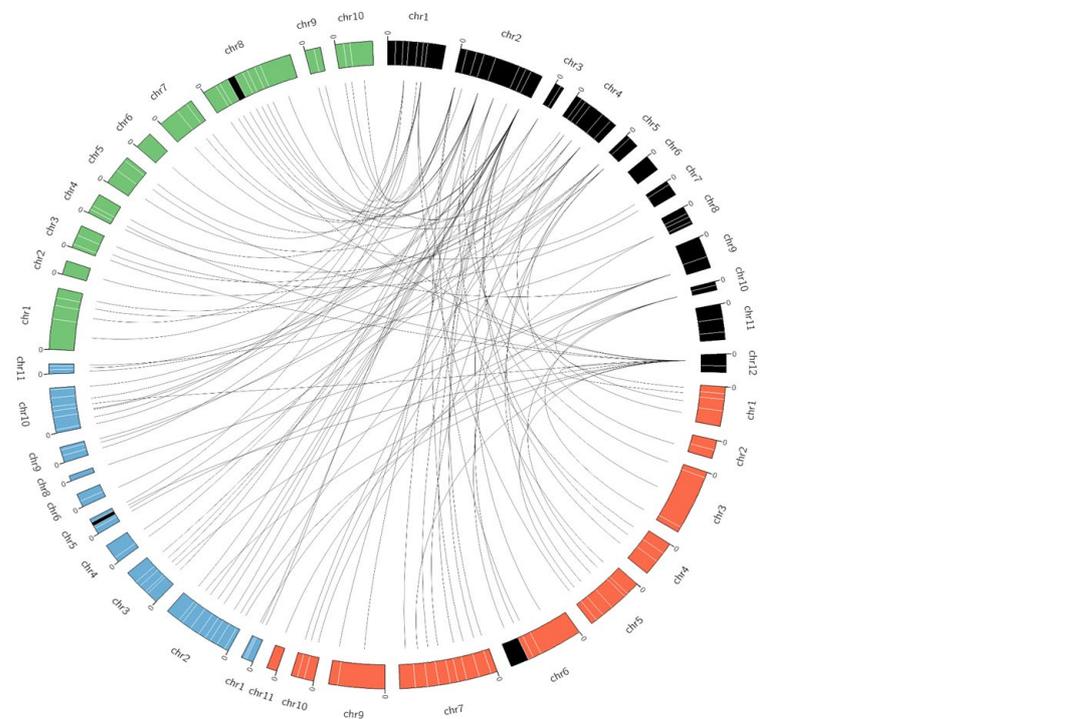


Figure 6. Circos map showing chromosomal position and collinearity of trihelix genes of cowpea, mung bean, adzuki bean and rice chromosomes. The red, green, blue and black represent chromosome of cowpea, mung bean, adzuki bean and rice, respectively. The black line shows connection of the respective orthologs of two species in *Vigna* and rice.

3.4. Expression Profile of Trihelix Genes in Different Tissue in Cowpea

The expression profile of major plant tissues such as leaf, root, stem, pod, seed and mix-tissue across developmental time series of cowpea were investigated to understand the expression profile of the trihelix genes in the cowpea. All 39 cowpea trihelix genes were found expressing in some or other studied tissue type (Figure 7). Most of the trihelix genes showed moderate expression and only a few genes showed a high level of expression. The high level of expression of most genes was found in seed tissues 18 days after pollination and the highest expression of genes (VungMSL36.1) was also observed in the same tissue. Detailed information on all the RNA-Seq data studies for the expression analysis is given in Table S1.

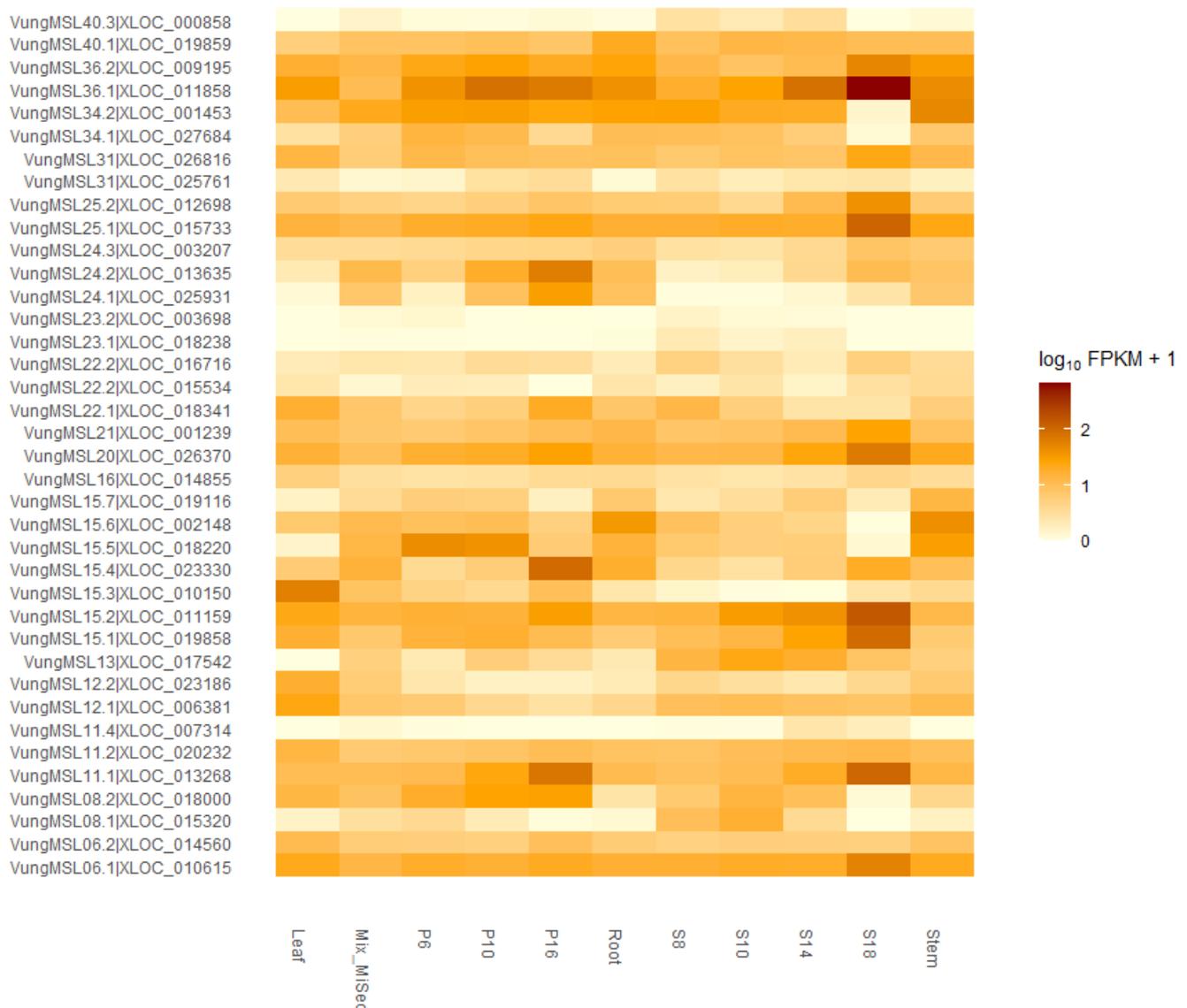


Figure 7. Expression of cowpea trihelix genes of major plant tissues at different developmental stages. (S8: Seed 8 days after pollination, S10: Seed 10 days after pollination, S14: Seed 14 days after pollination, S18: Seed 18 days after pollination, P6: Pod 6 days after pollination, P10: Pod 10 days after pollination, P16: Pod 16 days after pollination, Mix_MiSeq: Mix tissues). Heatmap was generated using cummerbund R package based on the FPKM (Fragments per kilobase of transcript per million mapped reads) value. Expression levels are depicted by same color with varying gradient from dark to light on the scale. Expression high: Expression value > 2 , Expression medium: Expression value > 1 , Expression low: $0 < \text{Expression value} < 1$.

3.5. Analysis of Rice Shattering 1 (Sha1) Orthologs in Vigna Species

To understand the extent of variation or conservation in the orthologs of rice SHA1 in four *Vigna* species, respective orthologs were studied by multiple sequence alignment with cultivated (OsMSL23) and wild rice (*O. rufipogon*). The cowpea, adzuki bean and rice bean each have two genes orthologous to OsMSL23. The mung bean does not seem to have a true ortholog of OsMSL23, while the nearest gene, VradMSL12, showed 36 and 35% similarity to the respective orthologs in cultivated and wild rice. Two genes of the cowpea, VungMSL23.1 and VungMSL23.2, showed 73 and 58% similarity with OsMSL23 and SHA1 in cultivated and wild rice, respectively. Adzuki bean and rice bean genes also showed comparable similarity with cultivated and wild rice orthologs.

The sequence alignment of SHA1/OsMSL23 orthologs in the three *Vigna* species cowpea (VungMSL23.1 and VungMSL23.2), adzuki bean (VangMSL23.1, VangMSL23.2) and rice bean (VumbMSL23.1, VumbMSL23.2) at the nucleotide level showed a wild rice variant base ‘Guanine’ as against ‘Thymine’ in cultivated rice (Figure 8a). At the amino acid level also, all the *Vigna* counterparts of SHA1/OsMSL23 showed a wild variant of amino acid ‘K’ as against ‘N’ in cultivated rice (Figure 8b). In the case of the mung bean, the closest gene, VradMSL12, did show a nucleotide variant (G) similar to wild rice; however, at the amino acid level, ‘R’ (Arginine) was present in place of ‘K’ in wild rice, which is also a positively charged amino acid as well as ‘K’.



Figure 8. Multiple sequence alignment of SHA1/Sh4 orthologs in *Vigna* crops (a) Alignment between the coding sequences (a) and amino acid sequences (b) of SHA1 of *O. rufipogon*, OsMSL23 of cultivated rice and respective orthologs in *Vigna*. The box with orange lines highlights single nucleotide (G to T) and amino acid (K to N) variations in respective crops.

Since a single amino acid change (Lysine to Asparagine) in OsMSL23 (Os04g57530.1) from the SH4 group of rice trihelix genes was found to have a substantial effect on gaining the non-shattering trait in rice [7], the effect of such a change in amino acid on the protein stability of the seven corresponding proteins (VungMSL23.2, VangMSL23.1, VangMSL23.2, VumbMSL23.1, VumbMSL23.2 and VradMSL12) of the four *Vigna* crops was studied. The 3-D protein structure of all the seven candidate proteins and their mutated variants was constructed by substituting amino acid (K to N) at the respective site (Figure 9). Both versions of each protein were compared on the basis of minimum stable energy and whether

the change in the amino acid had a stabilizing or destabilizing effect on the respective proteins. It was observed that both the cowpea proteins (VungMSL23.1, VungMSL23.2) and one protein each from the adzuki bean (VangMSL23.1), rice bean (VumbMSL23.1) and mung bean (VradMSL12) showed a significant reduction in the minimum energy of proteins after the change in amino acid from K to N, which had a stabilizing effect on these proteins (Table 5).

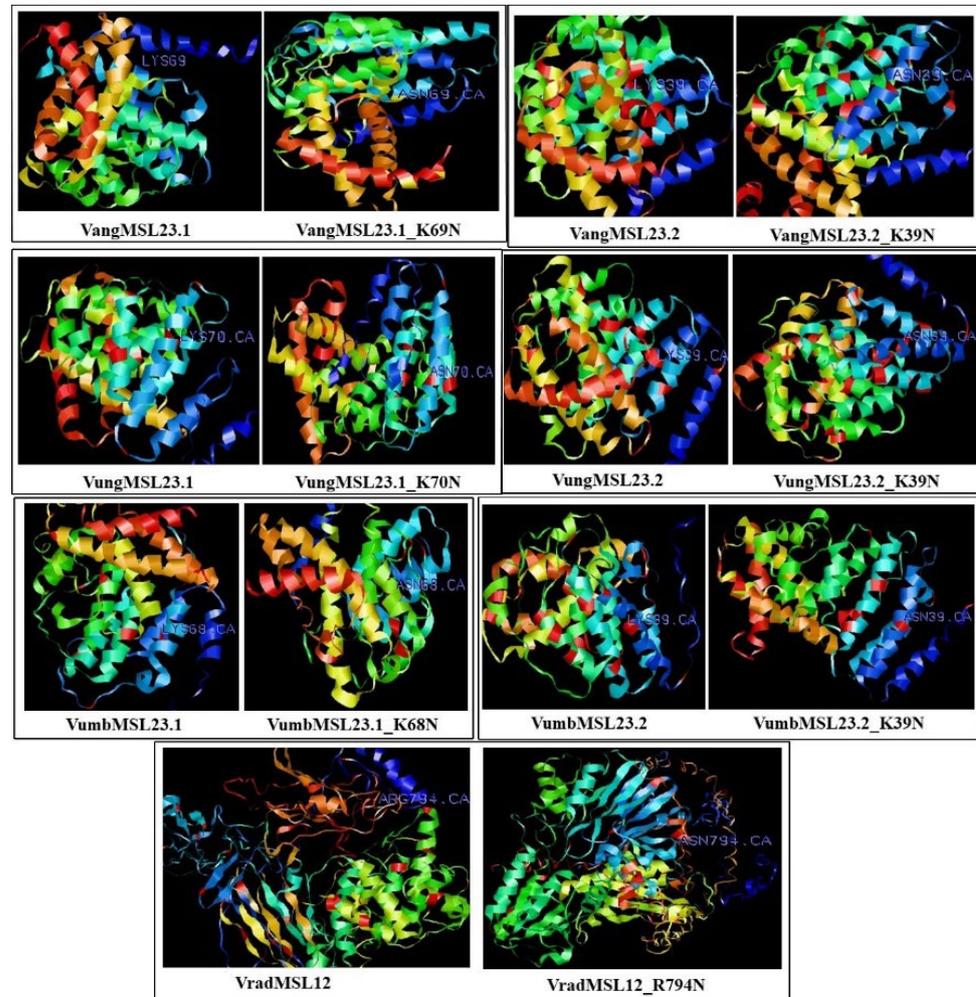


Figure 9. The 3D structure of proteins representing two variants of OsMSL23 homolog in *Vigna* crops showing two variants of each protein (K/R to N substitution).

Table 5. Minimum stable potential energy of candidate protein for shattering and its mutation variants.

Candidate Protein	Minimum Potential Energy	Mutation Variants of Candidate Protein	Minimum Potential Energy	Effect of Mutation
VungMSL23.1	−1,652,396.5	VungMSL23.1_K70N	−1,878,625.2	Stabilizing
VungMSL23.2	−1,605,236.2	VungMSL23.2_K39N	−1,676,144	Stabilizing
VangMSL23.1	−1,678,593.1	VangMSL23.1_K69N	−1,734,201.9	Stabilizing
VangMSL23.2	−1,567,263.5	VangMSL23.2_K39N	−1,546,699.9	Destabilizing
VumbMSL23.1	−1,555,491.4	VumbMSL23.1_K68N	−1,579,698.2	Stabilizing
VumbMSL23.2	−1,669,826.2	VumbMSL23.2_K39N	−1,653,073.4	Destabilizing
VradMSL12	−3,030,210	VradMSL12_R794N	−3,738,055.5	Stabilizing

4. Discussion

The *Vigna* group of crops, including the cowpea, mung bean, adzuki bean and rice bean, are not only economically important but are crucial for nutritional security and alleviating malnutrition in underdeveloped and developing countries. Unlike cereals, these crops have received limited attention from researchers and genomic intervention for genetic improvement. Recently, there have been noteworthy advances in terms of genome and transcriptome sequencing in these crops [30–33], which could facilitate comparative genomics to solicit benefit from the advances in other model crops. In these perspectives, here an important trihelix transcription factor family has been identified and characterized in four *Vigna* crops by making use of bioinformatic and comparative genomic approaches.

The total number of trihelix genes identified in the cowpea (39), mung bean (35), adzuki bean (41) and rice bean (50) was comparable to previous studies on rice (41), arabidopsis (28), tomato (36), chrysanthemum (20), *Brassica rapa* (52) and tartary buckwheat (31); however, the total number was less than in soybean (71) and wheat (94) [18–21,51–54]. In the rice bean, the number of trihelix genes identified in this study may change with the availability of the whole genome sequence in the public domain.

To facilitate unified nomenclature for easy comparison, the orthology-based nomenclature was proposed for trihelix genes in four *Vigna* crops with the basis of rice trihelix genes. The orthology-based nomenclature was earlier followed for the MAPK, MAPKK and CDPK gene families in rice, poplar and pigeon pea [55–57]. Such systematic nomenclature eases the species-to-species comparisons and helps in the functional characterization of homologs that show clear orthologous relationships with candidates from the model plant species. However, the conservation of biological functions within such relationships is to be resolved [56,57].

Previously, the phylogeny-based classification of the rice trihelix transcription factor family was given with three distinctive sub-families GT α , GT β and GT γ [58]. Further study on arabidopsis, rice, chrysanthemum, brassica and medicago further expanded the family into five sub-divisions GT-1, GT-2, GT γ , S1P1 and SH4 [19–21,51,59]. A new subfamily of trihelix, GT δ , was proposed in tomato and rice [18,21]. In the present study, the classification of the trihelix gene family in *Vigna* species was followed according to the classification in rice [21]. The predicted trihelix genes in *Vigna* species were classified into five subfamilies, GT, GT γ , S1P1, SH4 and GT δ , based on the evolutionary analysis with rice. In addition, all the predicted proteins of the trihelix family from *Vigna* species were predicted to be localized in the nucleus except OsMSL12 orthologs, which are found in the mitochondria and chloroplast (Tables 1–4), and VumbMSL16.3, present in cytoplasm. In congruence with this, in rice OsMSL12 was reported to be localized in chloroplast [21]. Transcription factors are known to be localized in the nucleus for their role in gene regulation. Recently, NAC102, an NAC family transcription factor, has been identified for the first time to directly regulate chloroplast gene expression [60]. It will be intriguing to see if the trihelix transcription factor particularly, an ortholog of OsMSL12 in *Vigna* species, has any role to play in chloroplastic gene regulation.

The exon numbers in the genes and gene structure were often found consistent for orthologous groups and for a given subgroup [21]. Further, with respect to the conserved motifs in all the trihelix proteins in *Vigna* species, a total 10 motifs were found to be distributed differentially, although mostly with similar patterns within a subfamily. Similar observations have also been reported in other crops, including soybean and rice (21,53). Mostly, the trihelix members from the same clade and particularly the same orthologs share a similar distribution of motif as well as exon\intron structure.

A fourth amphipathic α -helix with a strongly conserved motif ((F/Y)-(F/Y)-x-x-(L/I/M)-x-x-(L/I/M) followed by another three in the trihelix protein were previously shown to be located on the trihelix DNA binding domain of the GT-1, GT-2, GT γ and S1P1 clades in arabidopsis [13]. Almost all *Vigna* trihelix proteins contain this fourth helix with the same general sequences downstream of the other three helices on the DNA binding domain of the GT, S1P1 and GT γ clades (Figures S1, S2 and S4). This was not

present in the SH4 (Figure 4) and GT δ (Figure S3) clades; rather, the SH4 clade carries an extended third helix. This fourth helix is shown to be required for DNA binding [13]. The non-synonymous to synonymous substitution rate (dN/dS) for trihelix genes with their rice ortholog revealed that the ratio was significantly less than 1 for most of the trihelix genes (Tables S2–S5), indicating purifying or stabilizing selection in the course of evolution. Similar observations have been reported for other gene families in rice [61].

Domestication traits such as the loss/reduction of seed shattering or pod dehiscence were acquired in several crop plants independently by convergent phenotypic evolution [4]. Independent evolution for pod shattering in two closely related species, the common bean and cowpea, showed that a convergent domestication might take place when mutation occurs at orthologous loci [62]. Similarly, in arabidopsis and the common bean, AtMYB26 and its orthologue, PvMYB26, have been identified for pod shattering, which shows that orthologous genes preserved shattering function, and the associated pathways between and beyond closely related species [63]. MYB26 as a target for the non-shattering trait in the adzuki bean and yard-long bean has also been shown [64]. The genetic basis of domestication in many of the *Vigna* species is still unresolved and yield loss due to pod shattering is reported as a major concern [65,66].

In rice (*Oryza sativa*), the non-shattering trait was gained due to single nucleotide substitution from G to T at SHA1/Shattering 4 (Sh4), which encodes trihelix transcription factor Os04g57530.1 (OsMSL23) [7,8].

It was observed in our study that the ortholog of OsMSL23 in all three *Vigna* crops (cowpea, adzuki bean and rice bean) harbors the nucleotide (G) and amino acid variant 'K' as that of wild rice (Figure 8). In the mung bean, although a true ortholog of OsMSL23 is not present, the nearest gene, VradMSL12, also possesses the same wild variant at the nucleotide level (G), amino acid (R) however, which is of the same group as Lysine. It is interesting to know that *Vigna* crops display shattering to different degrees [65,66]. Besides the single amino acid variant, for the gene there is overall substantial sequence homology at the nucleotide and amino acid levels between *Vigna* crops and rice (Figure 8).

It will be interesting to see if the non-shattering trait can be gained in the four *Vigna* crops by attempting the same mutation as in cultivated rice. The expression study of the trihelix genes in the cowpea does show expression of *VungMSL23.1* and *VungMSL23.2* in seed tissues, albeit at a lower level (Figure 7).

In this direction, the 3-D protein structures of OsMSL23 orthologs in the cowpea (*VungMSL23.1* *VungMSL23.2*), adzuki bean (*VangMSL23.1*, *VangMSL23.2*), rice bean (*VumbMSL23.1*, *VumbMSL23.2*) and the nearest protein in mung bean (*VradMSL12*) with the 'K' and mutated 'N' amino acid variants at respective positions were constructed and assessed for stability. It was found that five of the seven proteins showed lower minimum energy than the native proteins and had a stabilizing effect (Figure 9 and Table 5). In the cowpea, *VungMSL23.1* showed better stability than *VungMSL23.2*. Accordingly, in the cowpea, *VungMSL23.1*, adzuki bean, *VangMSL23.1*, rice bean, *VumbMSL23.1* and in the mung bean, *VradMSL12* could form a potential target for genome editing for the non-shattering trait. It is, however, imperative to undertake the wet-lab validation and detailed characterization of the same. The effect of single amino acid change on protein stability has also been shown in other crops [67].

Recently, a genome editing technique has been suggested as a potential tool to target genes for the accelerating domestication of semi-domesticated or wild plants. It has been advocated that if the complete genome sequence of the desired plant is available then, based on the established domesticated genes, the target orthologous genes can be identified in that species for genome editing to accelerate domestication [68]. Indeed, recently 'De novo domestication' or 'Neo-domestication' has been proposed as a novel strategy for crop breeding for gaining the desired domestication traits [68–70].

Probably, the time clock has struck for 'De novo domestication' or 'Neo-domestication' for the accelerated and targeted introduction of domestication traits using genome editing in crop plants [68,71].

5. Conclusions

In this study, through genome-wide identification, the trihelix gene family comprising 39, 35, 41 and 50 genes have been identified in the cowpea, mung bean, adzuki bean and rice bean, respectively. The identified genes in the four *Vigna* crops have been characterized using bioinformatics and comparative genomic tools. The gene structure and conserved motifs in proteins were relatively consistent within the sub-groups of the gene family. The non-synonymous to synonymous substitution rate for trihelix genes in the four *Vigna* crops suggested purifying or stabilizing selection. Further, potential candidate genes were identified to target the non-shattering traits in the *Vigna* crops. The proposed change in the single nucleotide leading to the substitution of an amino acid showed a stabilizing effect at the protein level. It was proposed to apply the genome editing tools on the identified target trihelix genes in the four *Vigna* crops to have further insight into their role in the non-shattering phenotype, thereby accelerating domestication. The study gives an elaborate understanding of various aspects of the trihelix transcription factor family in four *Vigna* crops.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agriculture12122172/s1>, Figure S1: Sequence alignment of trihelix domain of GT group present in four *Vigna* species. Figure S2: Sequence alignment of trihelix domain of S1P1 group present in four *Vigna* species. Figure S3: Sequence alignment of trihelix domain of GT δ group present in four *Vigna* species. Figure S4: Sequence alignment of trihelix domain of GT γ group present in four *Vigna* species. Supplementary tables; Table S1: Detailed information RNASeq data of cowpea used for gene expression analysis of trihelix genes. Table S2: Non-Synonymous/Synonymous substitution rate for Trihelix genes of cowpea; Table S3: Non-Synonymous/Synonymous substitution rate for Trihelix genes of mung bean. Table S4: Non-Synonymous/Synonymous substitution rate for Trihelix genes of adzuki bean. Table S5: Non-Synonymous/Synonymous substitution rate for Trihelix genes of rice bean.

Author Contributions: Conceptualization, D.P.W. and S.A.; validation, S.K. and S.M.; formal analysis, S.K. and R.M.; investigation, S.K.; resources, S.A. and A.R.; writing—original draft preparation, S.K. and D.P.W.; writing—review and editing, D.P.W. and S.A.; supervision, S.A., A.R., S.J. and D.P.W.; funding acquisition, S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by ICAR National Fellow project.

Data Availability Statement: Not applicable.

Acknowledgments: During the period of study, Shweta Kumari was supported by a fellowship from ICAR-IARI, New Delhi. Sunil Archak was supported by an ICAR National fellowship. Computational facilities provided by ICAR-NBPGR and ICAR-IASRI are gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Rice bean Trihelix genes and protein sequences.

References

1. Ross-Ibarra, J.; Morrell, P.L.; Gaut, B.S. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (Suppl. S1), 8641–8648. [[CrossRef](#)] [[PubMed](#)]
2. Olsen, K.M.; Wendel, J.F. A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annu. Rev. Plant. Biol.* **2013**, *64*, 47–70. [[CrossRef](#)] [[PubMed](#)]
3. Dong, Y.; Wang, Y.Z. Seed shattering: From models to crops. *Front. Plant. Sci.* **2015**, *6*, 476. [[CrossRef](#)] [[PubMed](#)]
4. Vittori, V.D.; Gioia, T.; Rodriguez, M.; Bellucci, E.; Bitocchi, E.; Nanni, L.; Attene, G.; Rau, D.; Papa, R. Convergent evolution of the seed shattering trait. *Genes* **2019**, *10*, 68. [[CrossRef](#)]
5. Lin, Z.; Li, X.; Shannon, L.M.; Yeh, C.T.; Wang, M.L.; Bai, G.; Peng, Z.; Li, J.; Trick, H.N.; Clemente, T.E.; et al. Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* **2012**, *44*, 720–724. [[CrossRef](#)]
6. Li, C.; Zhou, A.; Sang, T. Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytol.* **2006**, *170*, 185–193. [[CrossRef](#)]

7. Li, C.; Zhou, A.; Sang, T. Rice domestication by reducing shattering. *Science* **2006**, *311*, 1936–1939. [[CrossRef](#)]
8. Lin, Z.; Griffith, M.E.; Li, X.; Zhu, Z.; Tan, L.; Fu, Y.; Zhang, W.; Wang, X.; Xie, D.; Sun, C. Origin of seed shattering in rice (*Oryza sativa* L.). *Planta* **2007**, *226*, 11–20. [[CrossRef](#)]
9. Zhou, D.X. Regulatory mechanism of plant gene transcription by GT-elements and GT-factors. *Trends Plant. Sci.* **1999**, *4*, 210–214. [[CrossRef](#)]
10. Qin, Y.; Ma, X.; Yu, G.; Wang, Q.; Wang, L.; Kong, L.; Kim, W.; Wang, H.W. Evolutionary history of trihelix family and their functional diversification. *DNA Res.* **2014**, *21*, 499–510. [[CrossRef](#)]
11. Nagano, Y. Several features of the GT-factor trihelix domain resemble those of the Myb DNA-binding domain. *Plant Physiol.* **2000**, *124*, 491–494. [[CrossRef](#)] [[PubMed](#)]
12. Riechmann, J.L.; Heard, J.; Martin, G.; Reuber, L.; Jiang, C.; Keddie, J.; Adam, L.; Pineda, O.; Ratcliffe, O.J.; Samaha, R.R.; et al. Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the *rbcS-3A* gene. *EMBO J.* **1987**, *6*, 2543–2549.
13. Kaplan-Levy, R.N.; Brewer, P.B.; Quon, T.; Smyth, D.R. The trihelix family of transcription factors—light, stress and development. *Trends Plant. Sci.* **2012**, *17*, 163–171. [[CrossRef](#)]
14. Gilmartin, P.M.; Memelink, J.; Hiratsuka, K.; Kay, S.A.; Chua, N.H. Characterization of a gene encoding a DNA binding protein with specificity for a light-responsive element. *Plant Cell.* **1992**, *4*, 839–849. [[PubMed](#)]
15. Kay, S.A.; Keith, B.; Shinzaki, K.; Chye, M.L.; Chua, N.H. The rice phytochrome gene: Structure, autoregulated expression, and binding of GT-1 to a conserved site in the 5' upstream region. *Plant Cell.* **1989**, *1*, 351–360. [[PubMed](#)]
16. Dehesh, K.; Bruce, W.B.; Quail, P.H. A trans-acting factor that binds to a GT-motif in a phytochrome gene promoter. *Science* **1990**, *250*, 1397–1399. [[CrossRef](#)] [[PubMed](#)]
17. Dehesh, K.; Hung, H.; Tepperman, J.M.; Quail, P.H. GT-2: A transcription factor with twin autonomous DNA-binding domains of closely related but different target sequence specificity. *EMBO J.* **1992**, *11*, 4131–4144. [[CrossRef](#)] [[PubMed](#)]
18. Yu, C.; Cai, X.; Ye, Z.; Li, H. Genome-wide identification and expression profiling analysis of trihelix gene family in tomato. *Biochem. Biophys. Res. Commun.* **2015**, *468*, 653–659. [[CrossRef](#)]
19. Yasmeen, E.; Riaz, M.; Sultan, S.; Azeem, F.; Abbas, A.; Riaz, K.; Ali, M.A. Genome-wide analysis of trihelix transcription factor gene family in *Arabidopsis thaliana*. *Pak. J. Agric. Sci.* **2016**, *53*.
20. Song, A.; Wu, D.; Fan, Q.; Tian, C.; Chen, S.; Guan, Z.; Xin, J.; Zhao, K.; Chen, F. Transcriptome-wide identification and expression profiling analysis of chrysanthemum trihelix transcription factors. *Int. J. Mol. Sci.* **2016**, *17*, 198. [[CrossRef](#)]
21. Li, J.; Zhang, M.; Sun, J.; Mao, X.; Wang, J.; Wang, J.; Liu, H.; Zheng, H.; Zhen, Z.; Zhao, H.; et al. Genome-Wide Characterization and Identification of Trihelix Transcription Factor and Expression Profiling in Response to Abiotic Stresses in Rice (*Oryza sativa* L.). *Int. J. Mol. Sci.* **2019**, *20*, 251. [[CrossRef](#)] [[PubMed](#)]
22. Brewer, P.B.; Howles, P.A.; Dorian, K.; Griffith, M.E.; Ishida, T.; Kaplan-Levy, R.N.; Kilinc, A.; Smyth, D.R. PETAL LOSS, a trihelix transcription factor gene, regulates perianth architecture in the *Arabidopsis* flower. *Development* **2004**, *131*, 4035–4045. [[CrossRef](#)] [[PubMed](#)]
23. Breuer, C.; Kawamura, A.; Ichikawa, T.; Tominaga-Wada, R.; Wada, T.; Kondou, Y.; Muto, S.; Matsui, M.; Sugimoto, K. The trihelix transcription factor GTL1 regulates ploidy-dependent cell growth in the *Arabidopsis* trichome. *Plant Cell* **2009**, *21*, 2307–2322. [[CrossRef](#)] [[PubMed](#)]
24. Gao, M.J.; Lydiate, D.J.; Li, X.; Lui, H.; Gjetvaj, B.; Hegedus, D.D.; Rozwadowski, K. Repression of seed maturation genes by a trihelix transcriptional repressor in *Arabidopsis* seedlings. *Plant Cell* **2009**, *21*, 54–71. [[CrossRef](#)] [[PubMed](#)]
25. Barr, M.S.; Willmann, M.R.; Jenik, P.D. Is there a role for trihelix transcription factors in embryo maturation? *Plant Signal. Behav.* **2012**, *7*, 205–209. [[CrossRef](#)] [[PubMed](#)]
26. Pratap, A.; Malviya, N.; Tomar, R.; Gupta, D.S.; Kumar, J. Vigna. In *Alien Gene Transfer in Crop Plants.*, 1st ed.; Pratap, A., Kumar, J., Eds.; Springer: New York, NY, USA, 2014; Volume 2, pp. 163–189.
27. Kang, Y.J.; Satyawan, D.; Shim, S.; Lee, T.; Lee, J.; Hwang, W.J.; Lee, S.H. Draft genome sequence of adzuki bean, *Vigna angularis*. *Sci. Rep.* **2015**, *5*. [[CrossRef](#)] [[PubMed](#)]
28. Kang, Y.J.; Kim, S.K.; Kim, M.Y.; Lestari, P.; Kim, K.H.; Ha, B.K.; Lee, S.H. Genome sequence of mung bean and insights into evolution within *Vigna* species. *Nat. Commun.* **2014**, *5*, 1–9. [[CrossRef](#)]
29. Lonardi, S.; Muñoz-Amatriaín, M.; Liang, Q.; Shu, S.; Wanamaker, S.I.; Lo, S.; Close, T.J. The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* **2019**, *98*, 767–782. [[CrossRef](#)]
30. Kaul, T.; Eswaran, M.; Thangaraj, A.; Meyyazhagan, A.; Nehra, M.; Raman, N.M.; Balamurali, B. Rice Bean (*Vigna umbellata*) draft genome sequence: Unravelling the late flowering and unpalatability related genomic resources for efficient domestication of this underutilized crop. *bioRxiv* **2019**, 816595. [[CrossRef](#)]
31. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2011**, *40*, D1178–D1186. [[CrossRef](#)]
32. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
33. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *J. Bioinform.* **2006**, *22*, 1658–1659. [[CrossRef](#)] [[PubMed](#)]

34. Cantarel, B.L.; Korf, I.; Robb, S.M.; Parra, G.; Ross, E.; Moore, B.; Yandell, M. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **2008**, *18*, 188–196. [[CrossRef](#)] [[PubMed](#)]
35. Stanke, M.; Tzvetkova, A.; Morgenstern, B. AUGUSTUS at EGASP: Using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **2006**, *7*, S11. [[CrossRef](#)] [[PubMed](#)]
36. Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217. [[CrossRef](#)] [[PubMed](#)]
37. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [[CrossRef](#)] [[PubMed](#)]
38. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [[CrossRef](#)] [[PubMed](#)]
39. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **2008**, *9*, 299–306. [[CrossRef](#)]
40. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [[CrossRef](#)]
41. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37* (Suppl. S2), W202–W208. [[CrossRef](#)]
42. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [[CrossRef](#)] [[PubMed](#)]
43. Yu, C.S.; Cheng, C.W.; Su, W.C.; Chang, K.C.; Huang, S.W.; Hwang, J.K.; Lu, C.H. CELLO2GO: A web server for protein subCELLular LOcalization prediction with functional gene ontology annotation. *PLoS ONE* **2014**, *9*, e99368. [[CrossRef](#)] [[PubMed](#)]
44. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [[CrossRef](#)] [[PubMed](#)]
45. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)] [[PubMed](#)]
46. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34* (Suppl. 2), W609–W612. [[CrossRef](#)]
47. Yao, S.; Jiang, C.; Huang, Z.; Torres-Jerez, I. The *Vigna unguiculata* Gene Expression Atlas (VuGEA) from de novo assembly and quantification of RNA-seq data provides insights into seed maturation mechanisms. *Plant J.* **2016**, *88*, 318–327. [[CrossRef](#)]
48. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [[CrossRef](#)]
49. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [[CrossRef](#)]
50. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25. [[CrossRef](#)]
51. Wang, W.; Wu, P.; Liu, T.; Ren, H.; Li, Y.; Hou, X. Genome-wide analysis and expression divergence of the trihelix family in *Brassica rapa*: Insight into the evolutionary patterns in plants. *Sci. Rep.* **2017**, *7*, 6463. [[CrossRef](#)]
52. Ma, Z.; Liu, M.; Sun, W.; Huang, L.; Wu, Q.; Bu, T.; Chen, H. Genome-wide identification and expression analysis of the trihelix transcription factor family in tartary buckwheat (*Fagopyrum tataricum*). *BMC Plant Biol.* **2019**, *19*, 344. [[CrossRef](#)] [[PubMed](#)]
53. Liu, W.; Zhang, Y.; Li, W.; Lin, Y.; Wang, C.; Xu, R.; Zhang, L. Genome-wide characterization and expression analysis of soybean trihelix gene family. *PeerJ* **2020**, *8*, 8753. [[CrossRef](#)] [[PubMed](#)]
54. Xiao, J.; Hu, R.; Gu, T.; Han, J.; Qiu, D. Genome-wide identification and expression profiling of trihelix gene family under abiotic stresses in wheat. *BMC Genom.* **2019**, *20*, 287. [[CrossRef](#)] [[PubMed](#)]
55. Hamel, L.P.; Nicole, M.C.; Sritubtim, S.; Morency, M.J.; Ellis, M.; Ehrling, J.; Beaudoin, N.; Barbazuk, B.; Klessig, D.; Lee, J.; et al. Ancient signals: Comparative genomics of plant MAPK and MAPKK gene families. *Trends Plant Sci.* **2006**, *11*, 192–198. [[CrossRef](#)] [[PubMed](#)]
56. Hamel, L.P.; Sheen, J.; Seguin, A. Ancient signals: Comparative genomics of green plant CDPKs. *Trends Plant Sci.* **2014**, *19*, 79–89. [[CrossRef](#)]
57. Wankhede, D.P.; Aravind, J.; Mishra, S.P. Identification of Genic SNPs from ESTs and Effect of Non-synonymous SNP on Proteins in Pigeonpea. *Proc. Natl. Acad. Sci. USA* **2019**, *89*, 595–603. [[CrossRef](#)]
58. Fang, Y.; Xie, K.; Hou, X.; Hu, H.; Xiong, L. Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. *Mol. Genet. Genom.* **2010**, *283*, 157–169. [[CrossRef](#)]
59. Liu, X.; Zhang, H.; Ma, L.; Wang, Z.; Wang, K. Genome-Wide Identification and Expression Profiling Analysis of the Trihelix Gene Family Under Abiotic Stresses in *Medicago truncatula*. *Genes* **2020**, *11*, 1389. [[CrossRef](#)]
60. Xin, K.; Pan, T.; Gao, S.; Yan, S.A. Transcription Factor Regulates Gene Expression in Chloroplasts. *Int. J. Mol. Sci.* **2021**, *22*, 6769. [[CrossRef](#)]
61. Yadav, S.K.; Santosh Kumar, V.V.; Verma, R.K. Genome-wide identification and characterization of ABA receptor PYL gene family in rice. *BMC Genom.* **2020**, *21*, 676. [[CrossRef](#)]

62. Rau, D.; Murgia, M.L.; Rodriguez, M.; Bitocchi, E.; Bellucci, E. Genomic dissection of pod shattering in common bean: Mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant J.* **2019**, *97*, 693–714. [[CrossRef](#)] [[PubMed](#)]
63. Vittori, D.V.; Bitocchi, E.; Rodriguez, M.; Alseekh, S.; Bellucci, E. Pod indehiscence in common bean is associated to the fine regulation of PvMYB26 and a non-functional abscission layer. *BioRxiv* **2020**, 021972. [[CrossRef](#)]
64. Takahashi, Y.; Kongjaimun, A.; Muto, C.; Kobayashi, Y.; Kumagai, M. Same locus for non-shattering seed pod in two independently domesticated legumes, *Vigna angularis* and *Vigna unguiculata*. *Front. Genet.* **2020**, *11*, 748. [[CrossRef](#)] [[PubMed](#)]
65. Isemura, T.; Kaga, A.; Konishi, S.; Ando, T.; Tomooka, N.; Han, O.K.; Vaughan, D.A. Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Ann. Bot.* **2007**, *100*, 1053–1071. [[CrossRef](#)] [[PubMed](#)]
66. Vairam, N.; Lavanya, S.A.; Vanniarajan, C. Screening for pod shattering in mutant population of mung bean (*Vigna radiata* (L.) Wilczek). *J. Appl. Nat. Sci.* **2017**, *9*, 1787–1791. [[CrossRef](#)]
67. Wankhede, D.P.; Kumari, M.; Richa, T.; Aravind, J.; Rajkumar, S. Genome wide identification and characterization of Calcium Dependent Protein Kinase gene family in *Cajanus cajan*. *J. Environ. Biol.* **2017**, *38*, 167. [[CrossRef](#)]
68. Osterberg, J.T.; Xiang, W.; Olsen, L.I.; Edenbrandt, A.K.; Vedel, S.E. Accelerating the domestication of new crops: Feasibility and approaches. *Trends Plant Sci.* **2017**, *22*, 373–384. [[CrossRef](#)]
69. Yu, H.; Li, J. Breeding future crops to feed the world through de novo domestication. *Nat Commun.* **2022**, *13*, 1171. [[CrossRef](#)]
70. Zsogon, A.; Cermak, T.; Voytas, D.; Peres, L.E. Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: Case study in tomato. *Plant Sci.* **2017**, *256*, 120–130. [[CrossRef](#)]
71. Piasias, M.T.; Bakala, H.S.; McAlvay, A.C.; Mabry, M.E.; Birchler, J.A.; Yang, B.; Pires, J.C. Prospects of Feral Crop De Novo Re-Domestication. *Plant Cell Physiol.* **2022**, *63*, 1641–1653. [[CrossRef](#)]